COMPUTATIONAL INEFFICIENCY OF NONPARAMETRIC TESTS FOR CONSISTENCY OF DATA WITH HOUSEHOLD CONSUMPTION

RAHUL DEB
DEPARTMENT OF ECONOMICS, YALE UNIVERSITY

ABSTRACT. Recently Cherchye et al. (2007a) provided a nonparametric characterization of a general collective model for household consumption which included externalities and public consumption. This characterization involved different necessary and sufficient conditions. We identify a special case of the household model for which their sufficiency conditions are also necessary. We show that verification of this sufficiency condition is NP Complete which provides justification for the parametric techniques used in the empirical literature.

1. INTRODUCTION

The collective household model introduced by Chiappori (1988,1992) was the first notable divergence from the unitary approach, which assumes each household has a single utility function, that is, it acts as a single decision maker. The household model differs from the unitary model in the fact that it allows for (possibly diverging) individual preferences amongst different members of the household. It merely assumes that the household behaves Pareto optimally, with the weight of each member's utility function providing a measure of their bargaining power. This model has been widely utilized and tested in the empirical literature. [1]

Browning and Chiappori (1998) provided a parametric characterization of the general collective model. Their model assumes that the aggregate consumption of the household is observed along with the prices of the commodities. Furthermore, they assume that the empirical economist cannot observe which goods are publicly or privately

Date: March 26, 2007.
Key words and phrases. Collective household models, intrahousehold allocation, revealed preference, nonparametric analysis, NP Completeness.
I would like to thank Donald Brown, Itzhak Gilboa and Maher Said for useful discussions and guidance. All errors are mine.

[1]A few influential papers that use the household model are Udry (1996), Dercon, S. and Krishnan, P. (2000) and Duflo and Udry (2004).

consumed and that individual consumptions are not observed. They also allow for externalities as well as altruism.

Recently, Cherchye et al. (2007a) provided a nonparametric characterization of the general collective model. They provide separate necessary and sufficient conditions for the observed household data to be consistent with two decision makers and then generalize these conditions to K decision makers in a supplement to the above paper (Cherchye et al. (2007b)). Finally they provide simple algorithms for testing both the necessary and sufficient conditions. These algorithms are exponential in the number in the number of observed data points although the authors provide ways to potentially improve efficiency of these algorithms in their paper, as well as in a companion paper (Cherchye et al. (2005)).

This paper identifies an important special case of the household model for which the sufficient conditions in Cherchye et al. (2007a) are also necessary. This is a straightforward generalization of the situation dependent dictatorship model in their paper, but it provides richer interpretations of data that satisfy their sufficiency conditions. It allows for a small number of unobserved changes in bargaining power of household members without requiring that each observation corresponds to a particular dictator.

The main result of the paper is that testing the nonparametric sufficiency conditions is a computationally difficult problem. In other words, we show that testing the sufficiency conditions is NP Complete: there is no known polynomial (in the number of data points) time algorithm which can verify whether an arbitrary set of data points satisfies the sufficiency conditions.

Although the main result is negative, it provides a justification for the existing practice of using a parametric framework in the empirical literature. Moreover it motivates the bridging of the gap between known necessary and sufficient conditions for consistency with the general collective household model. A more general sufficient condition would perhaps be easier to test.

2. The Model

We consider a K person household. An observed *data set* is a finite set of price - household consumption vectors $D = \{(p_i, x_i)\}_{i=1}^N$, where $(p_i, x_i) \in \mathbb{R}^l_{++} \times (\mathbb{R}^l_+ \backslash \{0\})$ and $1 \leq N < \infty$. All goods can be consumed privately, publicly or both. Thus an arbitrary consumption bundle x can be decomposed as

$$x = \sum_{k=1}^{K} x^k + x^h$$

where x^k is the (unobserved) private consumption of household member k and x^h is the (unobserved) public consumption of the household.

We consider general preferences such as those in Browning and Chiappori (1998) that depend not only on private and public consumption but also on other household members' private quantities. This allows for altruism and externalities. We restrict the case to positive externalities as in Cherchye et al. (2007a). This implies each household member k has a utility function $U^k(x^1, \ldots, x^K, x^h)$ which is nondecreasing in all its arguments. We also assume U^k is locally non-satiated.

For each aggregate observation (p_i, x_i) we define *feasible personalized quantities* $\hat{x}_i$ as

$$\hat{x}_i = (x_i^1, \ldots, x_i^K, x_i^h) \in \mathbb{R}^{(K+1)l}, \qquad \text{with} \qquad x_i = \sum_{k=1}^{K} x_i^k + x_i^h$$

This decomposition captures feasible individual and public consumptions from the given observed data. Given this we can now define the condition for collective rationalization of data D.

Definition 1 (Collective Rationalization)**.** *Let $D = \{(p_i, x_i)\}_{i=1}^N$ be an observed household consumption data set. Utility functions $U^1, U^2, \ldots, U^K$ provide a* **collective rationalization** *of D if for each observation i there exist feasible individual and public consumptions $\hat{x}_i = (x_i^1, \ldots, x_i^K, x_i^h)$ and Pareto weights $(\mu_i^1, \ldots, \mu_i^K) \in \mathbb{R}^K_+$ such that*

$$\sum_{k=1}^{K} \mu_i^k U^k(\hat{x}_i) \geq \sum_{k=1}^{K} \mu_i^k U^k(\hat{z})$$

for all $\hat{z} = (z^1, \dots, z^K, z^h) \in \mathbb{R}_+^{(K+1)l}$ *with* $p_i'(z^1 + \dots + z^K + z^h) \leq p_i' x_i$ *and* $\sum_{k=1}^K \mu_i^k = 1$.

We now define a special case of the above model

Definition 2 (Consistency with K Decision Makers). *Let* $D = \{(p_i, x_i)\}_{i=1}^N$ *be an observed household consumption data set. We say the data* D *is* **consistent with** K **decision makers** *if there exist utility functions* $U^1, U^2, \dots, U^K$ *and feasible individual and public consumptions* $\hat{x}_i = (x_i^1, \dots, x_i^K, x_i^h)$ *and Pareto weights* $(\mu_i^1, \dots, \mu_i^K) \in M \subset \mathbb{R}_+^K$ *such that* $|M| \leq K$ *and*

$$\sum_{k=1}^K \mu_i^k U^k(\hat{x}_i) \geq \sum_{k=1}^K \mu_i^k U^k(\hat{z})$$

for all $\hat{z} = (z^1, \dots, z^K, z^h) \in \mathbb{R}_+^{(K+1)l}$ *with* $p_i'(z^1 + \dots + z^K + z^h) \leq p_i' x_i$ *and* $\sum_{k=1}^K \mu_i^k = 1$.

This special case only allows situations where the bargaining power between household members doesn't change more times than the number of decision makers. This is clearly still a generalization of the unitary model as the data can potentially be rationalized by a single household utility function.

This special case includes the *situation dependent dictatorship* model in Cherchye et al. (2007a) (in fact, it is easy to show that they are the same (Theorem 1)). As an example, consider supermarket purchase data of a two person household, where the identity of the household member who went shopping at each observation is unobserved. Thus the person whose turn it is to do the chores gets to act as a dictator and make decisions on the behalf of the entire household.

This special case also takes into account situations where there are a small number of exogenous unobserved shocks, which result in a change of bargaining power. Consider a household consisting of a happily married couple. The wife and husband have an understanding and hence consume according to a some fixed bargaining parameters $1 - \mu$ and μ respectively. The only strife in the household occurs when the wife's mother pays a visit. Then in exchange for being sociable, the husband has more bargaining power $\mu' > \mu$ (the wife's bargaining power becoming $1 - \mu'$), hence consuming more of the goods that he prefers. An example of a model in the empirical literature which considers such a one time change in bargaining power is Oreffice (2005).

Finally we define the Generalized Axiom of Revealed Preference (GARP) which will be used in the proofs of our results.

Definition 3 (GARP). *Given arbitrary data set* $D = \{(p_i, x_i)\}_{i=1}^N$. *For any two consumption bundles* x_1 *and* x_2 *we say* $x_1 \succ_{R^0} x_2$ *if* $p_1'x_1 \geq p_1'x_2$. *We say* $x_1 \succ_P x_2$ *if* $p_1'x_1 > p_1'x_2$. *Finally we say* $x_1 \succ_R x_2$ *if for some sequence of observations* $(x_i, x_j, \ldots, x_m)$ *we have* $x_1 \succ_{R^0} x_i$, $x_i \succ_{R^0} x_j$, $\ldots, x_m \succ_{R^0} x_2$. *In other words relation* $\succ_R$ *is the transitive closure of* $\succ_{R_0}$. *The data* D *satisfies* **GARP** *if*

$$x_i \succ_R x_j \implies x_j \not\succ_P x_i \qquad \forall x_i, x_j$$

3. Results

We now show that the sufficiency conditions of Cherchye et al. (2007a) are also necessary for a data set D to be consistent with K decision makers.

Theorem 1. *The data set* $D = \{(p_i, x_i)\}_{i=1}^N$ *is consistent with* $K \leq N$ *decision makers if, and only if, we can partition the set* D *into disjoint sets* $D_1, D_2, \ldots, D_K$ *such that each* D_k *(*$1 \leq k \leq K$*) satisfies GARP.*

Proof. The proof is immediate from Definition 2. Since the situation dependent dictatorship model is consistent with K decision makers the sufficiency is immediately established from Cherchye et al. (2007a). The necessity follows from the following. The data points corresponding to particular fixed bargaining parameters will satisfy GARP, as each household member's utility functions is non decreasing. Since we restrict the number of distinct bargaining parameters to be less than K the necessity is obvious. □

Our main result establishes the impracticality of using nonparametric tests for consistency of observed data with the household model.

Theorem 2. *Given a fixed integer* $1 < K \leq N$. *Testing the consistency of the data* $D = \{(p_i, x_i)\}_{i=1}^N$ *with* K *decision makers is NP Complete.*

Proof. We will use the partition into forests problem to show the result. This is a known NP Complete problem (Garey and Johnson (1979)). We will proceed by mapping every instance of the partition into forests problem to a particular instance of the data consistency problem.

Consider a directed graph $G(V, E)$ and a given fixed positive integer $K \leq |V|$, where V is the set of vertices and E is the set of edges. We say there is a directed edge from vertex i to j if $(i, j) \in E$. Finally we define the out degree of a vertex i as the number of outward edges from the vertex i. The partition into forests problem is then defined as follows.

Given an arbitrary integer K, can the nodes of the graph be partitioned into $s \leq K$ disjoint sets $V_1, V_2, \ldots, V_s$ in such a way that for each V_j $(1 \leq j \leq s)$ the subgraph $G_j(V_j, E_j)$ induced by V_j contains no cycles; that is, the set of subgraphs G_j is a forest (set of trees)?

We start with an arbitrary directed graph $G(V, E)$ where $|V| = N$. We now construct an appropriate data set $\{(p_i, x_i)\}_{i=1}^N$ (in polynomial time) such that the graph induced by the relation $\succ_P$ on the data is the same as G. We proceed as follows.

Let us assume that the number of goods in each bundle is N^2. Each good is labeled by two numbers mn where $1 \leq m, n \leq N$. In a slight abuse of notation, we represent the mn^{th} good in the i^{th} observation by x_i^{mn}, and the price of the mn^{th} good in the i^{th} observation by p_i^{mn}, where $1 \leq i, m, n \leq N$.

We now construct the data set $D = \{(p_i, x_i)\}_{i=1}^N$, where price consumption bundle i corresponds to vertex $i \in V$, as follows

$$
\begin{array}{llll}
p_i^{ii} = 1 & x_i^{ii} = 1 & & \\
p_i^{ij} = 1 & x_i^{ij} = 1 & & \text{if } j \neq i \text{ and } (i,j) \in E \\
p_i^{ji} = 0 & x_i^{ji} = 0 & & \text{if } j \neq i \text{ and } (j,i) \in E \\
p_i^{ij} = 1 & x_i^{ij} = 0 & & \text{if } j \neq i \text{ and } (i,j) \notin E \\
p_i^{ji} = 0 & x_i^{ji} = \text{out degree of } (j) + 2 & & \text{if } j \neq i \text{ and } (j,i) \notin E \\
p_i^{kl} = 0 & x_i^{kl} = 0 & & \text{if } k, l \neq i
\end{array}
\tag{1}
$$

Clearly this data set can be constructed in polynomial time. The intuition behind the construction is as follows. Each observation i corresponds to a vertex. Each good corresponds to an edge of the graph, except goods of the form ii which ensure each observation reflects positive wealth. For any two observations i and j we select appropriate prices and consumption bundles for the goods ij and ji (the remaining goods not coming into play) so that the relation $\succ_P$ reflects whether there is an edge (i,j) or (j,i) in the graph G.

We now check to see if the graph induced by the relation $\succ_P$ on the above data set is the same as G. For an arbitrary $i \neq j$ and $i, j \in V$, if we have $(i,j) \in E$ then

$$\begin{aligned} &p_i' x_i = \text{out degree of } (i) + 1 \qquad \text{[which is at least 1]} \\ &p_i' x_j = 0 \\ &\implies p_i' x_i > p_i' x_j \\ &\implies x_i \succ_P x_j \end{aligned}$$

Similarly if we have $(i,j) \notin E$ then

$$\begin{aligned} &p_i' x_i = \text{out degree of } (i) + 1 \\ &p_i' x_j = \text{out degree of } (i) + 2 \\ &\implies p_i' x_i < p_i' x_j \\ &\implies x_i \not\succ_P x_j \end{aligned}$$

In the data set defined above, we have $x_i \succ_P x_j$ (equivalently $x_i \succ_{R^0} x_j$) if, and only if there is an edge from vertex i to vertex j in G (or $(i,j) \in E$). Hence the graph induced by the relation $\succ_P$ on the above data set is the same as G. Moreover, any subset $D' \subseteq D$ of the data set satisfies GARP if, and only if, the subgraph induced by the vertices corresponding to the data points in D' is acyclic. Hence, we can only solve the partition into forests problem for arbitrary graph G and fixed integer K, if we can check the consistency of the data set D with K decision makers.

The proof is not complete because our model does not allow data with zero prices or bundles where nothing is consumed. To solve this problem we can replace every instance of 0 by a very small ϵ in equation (1). We can once again verify the inequalities for each i. If $i \neq j$, $i, j \in V$ and $(i,j) \in E$

$$p_i'x_i > \text{out degree of}(i) + 1$$
$$p_i'x_j < N^2(N+2)\epsilon \qquad \text{[Because each vertex can have out degree at most } N\text{]}$$

If $(i,j) \notin E$

$$p_i'x_i < \text{out degree of}(i) + 1 + N^2(N+2)\epsilon$$
$$p_i'x_j > \text{out degree of}(i) + 2$$

For small enough $\epsilon < \frac{1}{N^2(N+2)}$ we will have $p_i'x_i > p_i'x_j$ when $(i,j) \in E$ and $p_i'x_i < p_i'x_j$ when $(i,j) \notin E$.

Hence, we can map (in polynomial time) any arbitrary graph G to a particular data set D such that for a fixed integer K we can solve the partition into forests problem for G only if, the data set D is consistent with K decision makers. Thus testing an arbitrary data set for consistency with K decision makers is NP Complete. □

4. Conclusions

The computational efficiency of tests are vital to the empirical economist. The nonparametric tests for consistency of data with the household model are clearly going to be difficult for the empiricist to implement. Hence, parametric tests using a particular functional form might be more informative.

However, there is scope to improve the nonparametric tests. A single, testable, necessary and sufficient condition for the consistency of data with the general household model, would allow us to test for a larger range of possibilities and could prove to be more computationally efficient. Another approach could be to sample (if necessary resample) a fixed number of points (a function of the number of decision makers K, we are testing for) each time and test this subset of the data for consistency. This will allow a compromise between the accuracy of the test and the computation time. Such approximate nonparametric tests could potentially be more informative than the existing parametric tests.

References

[1] Browning, M., and Chiappori, P. A. (1998), Efficient Intra-Household Allocations: A General Characterization and Empirical Tests. *Econometrica*, 66, 1241-1278.

[2] Cherchye, L., B. De Rock, and Vermuelen, F. (2005), Opening the Black Box of Intra- Household Decision-Making: Theory and Non-Parametric Empirical Tests of General Collective Consumption Models. *CentER Discussion Paper 2005-51*, Tilburg University [573].

[3] Cherchye, L., B. De Rock, and Vermuelen, F. (2007a), The Collective Model of Household Consumption: A Nonparametric Characterization. *Econometrica*, Vol. 75, No. 2 , 553-574.

[4] Cherchye, L., B. De Rock, and Vermuelen, F. (2007b), Supplement to The Collective Model of Household Consumption: A Nonparametric Characterization. *Econometrica Supplementary Material*, 75, http://www. econometricsociety.org/ecta/supmat/5499extensions.pdf. [554,565]

[5] Chiappori, P. A. (1988), Rational Household Labor Supply. *Econometrica*, 56, 63-90.

[6] Chiappori, P. A.(1992), Collective Labor Supply and Welfare. *Journal of Political Economy*, 100, 437-467.

[7] Dercon, S. and Krishnan, P. (2000). In Sickness and in Health: Risk Sharing within Households in Rural Ethiopia. *Journal of Political Economy*, Vol. 108 (4), 688–727.

[8] Duflo, E. and Udry, C. (2004). Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts and Consumption Choices *NBER Working Papers* 10498, National Bureau of Economic Research, Inc.

[9] Garey, M. and Johnson,D. (1979). Computers and Intractability - A Guide to the Theory of NP-completeness. *Freeman.*

[10] Oreffice, S. (2005). Did The Legalization of Abortion Increase Womens Household Bargaining Power? Evidence from Labor Supply. *Mimeo*, Clemson University.

[11] Udry, C. (1996). Gender, Agricultural Production, and the Theory of the Household. *Journal of Political Economy*, Vol. 104(5), 1010–1046.